

**UNCLASSIFIED**  
**AD 405 561**  

---

**DEFENSE DOCUMENTATION CENTER**  
**FOR**  
**SCIENTIFIC AND TECHNICAL INFORMATION**  
**CAMERON STATION, ALEXANDRIA, VIRGINIA**



**UNCLASSIFIED**

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

63-3-5

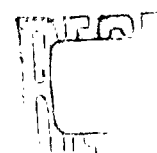
405561

405 561

MEMORANDUM  
RM-3625-PR  
MAY 1963

THE POSITION OF  
PREPOSITIONAL PHRASES IN RUSSIAN

Kenneth E. Harper



PREPARED FOR:  
UNITED STATES AIR FORCE PROJECT RAND

---

The **RAND** Corporation  
SANTA MONICA • CALIFORNIA

---

**MEMORANDUM**

**RM-3625-PR**

**MAY 1963**

**THE POSITION OF  
PREPOSITIONAL PHRASES IN RUSSIAN**

**Kenneth E. Harper**

This research is sponsored by the United States Air Force under Project RAND—contract No. AF 49(638)-700 monitored by the Directorate of Development Planning, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this Memorandum should not be interpreted as representing the official opinion or policy of the United States Air Force.

---

*The* **RAND** *Corporation*

1700 MAIN ST • SANTA MONICA • CALIFORNIA

---

PREFACE

An important objective in automatic language data-processing research is the development of computer programs for parsing written texts. These programs are useful both in language research and in such applications as machine translation, indexing, and abstracting. The present Memorandum is one of a series of RAND studies on sentence-structure description in Russian texts. The immediate objective is to improve the existing capability for Russian-English automatic translation.

SUMMARY

A problem frequently encountered in the automatic parsing of Russian texts is the correct structuring of prepositional phrases in sentences. Studies of text samples indicate that, when other criteria are absent, the syntactic governors of prepositions can be determined with a high degree of accuracy by reference to the relative position and part-of-speech of elements in the clausal environment.

CONTENTS

PREFACE .....	111
SUMMARY .....	v
Section	
1. INTRODUCTION .....	1
2. THE PREPOSITION IN STRUCTURE .....	3
3. THE RELATIVE POSITION OF A PREPOSITIONAL PHRASE AND ITS GOVERNOR .....	8
4. CONCLUSIONS .....	15
REFERENCES .....	17

## THE POSITION OF PREPOSITIONAL PHRASES IN RUSSIAN

### 1. INTRODUCTION

One of the primary goals of computational linguistics is the development of automatic parsing programs for use in processing written texts. There is an enormous utility for computer programs that will produce structural descriptions of sentences comparable to the descriptions produced by humans. Although both products are admittedly imperfect, given the present inadequacies of grammatical theory, the information generated in the course of automatic syntactic analysis is of immediate use in language study: the parsing programs themselves can be improved, and a "data base" is provided for testing the theoretical principles underlying the program. The parsing routine is a research tool for the automatic assembling of facts about the combinatorial properties of sentence elements; in particular, it is a means of achieving specificity in syntactic description. (In addition, automatic parsing has practical applications in such activities as machine translation, indexing, and abstracting.)

A parsing program must be based on a model of language, however imperfect and tentative that model may be. The program described in the present paper is based on a simple dependency grammar, adopted for linguistic research in Russian at The RAND Corporation [1]. In this model, the structure of a sentence is conceived as a tree-like set of relations among the words in the sentence. One word in every clause is said to be independent (in our convention, the predicate); except for this item, every other word in the clause "depends" on one and only one other word. (Double dependency is allowable in special instances, e.g., with relative pronouns.) The word on which a word depends is said to be its "governor"; the



latter term is used merely as a complement to "dependent," and does not necessarily correspond to the usage of the term in traditional grammatical description. The syntactic relationships designated here by dependency include instances of agreement and government (normally characterized by flexion in Russian), and complementation or modification of meaning. This model results in unique parsings for most sentences, assuming the acceptance of certain conventions. The latter are useful when the dependency of a word on a group of words is indicated; here, it is necessary to select as governor one word that will represent the group.

The automatic parsing program for Russian, as developed at RAND, has been described elsewhere [2]. For present purposes, we note only that a sample of some 10,000 sentences of text from Russian physics journals has been subjected to the program; the resulting descriptions have been verified and corrected by humans. All dependency relations between the word pairs in this text sample are recorded on magnetic tape. Automatic retrieval programs applied to this "processed" text enable researchers to conduct distributional studies (e.g., through concordances for specified words or syntactic constructions).

The present study deals with a common but difficult problem in machine parsing: the automatic structuring of prepositional phrases in sentences. As with other sentence elements, the preposition is said to depend on one other word (its governor); in turn, it governs at least one other word (its dependent, or object). The difficulty lies in the assignment of the correct governor for the preposition. We should stress the fact that we are not seeking to determine the governor of the prepositional phrase. Machine parsing proceeds by a comparison of the respective morphological and syntactic properties of pairs of word-tokens in the sentence; these properties are stored in the machine dictionary, and unless we create an

unmanageably large dictionary by storing prepositional phrases, we can parse only in terms of word pairs (preposition/governor and preposition/dependent). Can we construct a computer program capable of accounting for the numerous relationships that a preposition bears to its syntactic governor? Can we ignore, or can we utilize, the "semantic" information contained in the prepositional phrase,--information that the human parser takes full advantage of? The following discussion attempts to deal both with these theoretical problems and with the immediate, "practical" problem of improving the machine parsing program.

The main discussion will be prefaced by some remarks on the structuring of prepositional phrases in our version of dependency grammar.

## 2. THE PREPOSITION IN STRUCTURE

Since prepositions are relational words, it is convenient to think of a three-term structure: G/P/D (governor, preposition, dependent). (Other word-classes involved in three-term structures are coordinating and subordinating conjunctions, relative pronouns, and relative adverbs.) The relation of P to D can be specified morphologically in Russian, and presents no serious problem for the machine parser. Both word order and case of the D (assuming flexion) are precise and obligatory. In the computer program the task is performed by a simple matching of pairs of morphological codes: the dependent is said to be the first following occurrence whose codes match the part-of-speech and case requirement codes of the P. In rare instances, the machine is confronted with two possible dependents for the P. These exceptions are of two main types: those in which a following occurrence is homographic (e.g., IZ ETOGO TOCHNOGO OPREDELENIYA NE POLUCHAEM - "from this (an) exact definition we do not obtain"), and

those in which a nested structure is present (ETO ZAVISIT OT OPREDELYAEMOGO S POMOSHCH'YU SPEKTROMETRA ZNACHENIYA - "this depends on the obtained-with-the-aid-of-a-spectrometer value"). (In these examples, the P and the two possible Ds are underlined.) Here, the mechanical resolution of the problem depends on the correct structural description given to the whole clause.

The situation with respect to the preposition and its governor is far more complicated. The relative position of the two items is not constant, and morphology per se is of no help. Traditionally, Russian grammarians have used two terms to describe this relation: government and adjoinment (primykanie). It is widely agreed that there are substantial differences in the strength of the connection between P and G. Some grammarians have tried to distinguish between "strong" and "weak" government. A. M. Peshkovskij, for example, described the weakly governed preposition as one for which the connection may depend on such factors as word order or meaning; sometimes this situation creates ambiguity of meaning, and sometimes the prepositional phrase may be connected with the whole clause [3]. Strong government is said to contrast in the above respects, although the strength of the connection may vary considerably in different word combinations. In the case of adjoinment, the connection is felt to be weaker still, as with adverbial modification.

The validity of these distinctions has been seriously questioned. For example, the Academy of Sciences' Grammar of the Russian Language characterizes weak government as a "diffuse metaphor," and stresses the need for further study and greater precision in this area of grammar [4]. We could not agree more. From our point of view, strong and weak government, and adjoinment, represent little more than intuitive judgements of the kind of syntactic connection that has already been made. The gradations in

the strength of the connection are of little utility in analysis itself, since our immediate concern is to make the connection (in our terms, to find the G), not to assess its "quality." A separate treatment of this whole problem is planned. For the present, we remark only (i) that we propose to identify all connections, rejecting any implication that certain types of connections are less "important" than others, and (ii) that the frequency of occurrence in given texts of G/P word-pairs is a useful criterion in automatic parsing.

The criterion of frequency of G/P pairs, is, in fact, a test for the intuitive judgement that strongly governed complements answer questions implied or engendered by the governor [5]. If the verbs UEKHAT' = "to leave," or OTNOSIT' = "to relate," so to speak require complementation of the kind "whence," or "to what, or to whom," then we should frequently encounter these complements in written text. We have, in fact, retrieved from our physics text a large number of frequently occurring G/P pairs. Syntactic codes have been assigned in the machine dictionary to both members of such pairs; if, under appropriate conditions, the codes for these two words are matched during the

- automatic parsing routine, a pairing is effected. (The appropriate conditions include proper case for the object of the P, and the existence of previously established dependency pairs in the sentence, i.e., "precedence" [2]). The term "strong government" will be used in the present discussion to include the following types of G/P pairs:

- (i) pairs for which the P is an obligatory complement (VLIYAT' NA = "to influence," ZAVISET' OT = "to depend on");
- (ii) pairs for which the P is a frequent complement (SRAVNIT' S = "to compare with"; ZAVISIMOST' OT = "dependence on"); (iii) pairs in which the translation of the preposition can be effected only by its association with a given G. Pairs of the latter type are listed in a

separate paper [6]; here, the quality or strength of connection between G and P varies considerably (cf. BLAGODARIT' KOGO-TO ZA = "to thank someone for," POPRAVKA NA = "correction for," VEROYATNOST' OT = "probability of," RASPREDELENIE PO = "distribution with respect to," SUMMIROVAT' PO = "to sum over," RESHIT' CHEREZ = "to solve in terms of").

The current parsing routine operates on the principle that two members of a pre-assigned GP pair will be joined in a dependency pair if they co-occur in the same clause (under certain conditions). Our experience has been that the resulting syntactic analysis is correct with very few exceptions. The exceptions occur chiefly in structures that are essentially ambiguous in the immediate context. Thus, in TURISTY UEKHALI IZ MOSKVI = "the tourists left Moscow," the strong governor of IZ is the verb; the situation becomes ambiguous, however, with TURISTY IZ MOSKVI UEKHALI = "the tourists from Moscow left," or with UEZD TURISTOV IZ MOSKVI = "the departure of the tourists from Moscow." Since, in such cases, the ambiguity can be resolved only by reference to a larger context, we can only recognize that the co-occurrence of members of a G/P pair is not a guarantee that the connection can be correctly established.

In running text, the ratio of strongly governed Ps to all occurrences of Ps is rather low; in our physics text, the ratio is estimated at 1 to 5 for approximately 34,000 occurrences of Ps. Quantitatively, the major task is the attachment of weakly governed or "adjoined" prepositional phrases to the correct sentence element. In this case, there is no possibility of matching codes for G/P pairs.

So far as the human parser is concerned, three general situations may be noted.

- (1) The relation of P to G is clear, or can be

specified with a high degree of probability (ON UVIDEL KNIGU NA STOLE = "he saw the book on the table").

(ii) The relation of P to G is ambiguous. This situation is commonly found in the frame, transitive verb/noun object/prepositional phrase, where the latter can logically refer to (depend on) either the verb or the noun: ON NAPISAL SLOVA NA DOSKE = "He wrote the words on the blackboard." In some instances, the ambiguity may be irrelevant in translation ("I met the man on the corner"). In others, the structural description will affect the translation: "I hit the man for Nixon," "I hit the man with the ax," "I read the letter to John." In essence, this problem cannot be solved within the micro-context, although the probabilities may vary for different structuring.

(iii) The relation of P to G is not specific and is not relevant to meaning. Here, the relative position of the prepositional phrase is often the key to structural description; a shift in position, however, reflects only a shift in emphasis. For example, in the following sentences, the prepositional phrase would probably be connected with the preceding noun: "The temperature in the room rose," "The value for x was determined." Intuitively, we may doubt that any essential change in meaning results if the phrase is attached to the verb; the latter structuring is possible in the above examples, and is probable when the phrase follows the verb: "The temperature rose in the room," "The value was determined for x." The point is simply that in some structures the relation of G to P is less obvious and less dependent on meaning than in others.

In dealing with the first and last of these three situations, it is obvious that the human parser can bring to bear an enormous store of knowledge of the appropriate (i.e., the probable). It is precisely this kind of

"semantic" information that the machine lacks. One despairs of embedding in a machine program the information necessary to parse correctly sentences of the first type ("I read the book in bed," "I read the book under the dictionary," etc.), or to decide that the structuring is nonessential to meaning. If this is true, should we not admit that accurate machine parsing is an impossible goal? In the following, an estimate is given of the magnitude of the problem in running text; a solution is suggested in terms of the preposition relative to other sentence members.

### 3. THE RELATIVE POSITION OF A PREPOSITIONAL PHRASE AND ITS GOVERNOR

The relative ordering of sentence elements is an important factor in the redundancy of natural language. Although word order is freer in Russian than in English, there are a number of severe restrictions common to both. In many three-term structures, the relative ordering of syntactic elements is fixed. Thus, when two elements (A and B) are joined by a subordinate conjunction or a relative adverb (J), a maximum of two orderings are permitted: A/J/B ("I know/that/he is coming," "I know/where/he lives"), or, for special emphasis, J/B/A ("That/he is coming/ I know," "Where/he lives/I know"). The other four orderings of these elements are impossible, e.g., A/B/J ("I know/he is coming/that," "I know/he lives/where"). The restriction is simply that the order J/B be preserved. In the case of coordinate conjunctions and relative pronouns, only one ordering is found: A/J/B (cf. "and/men/women," "The man/I saw/whom"). It could be argued that these restrictions are due chiefly to convention, since they are not necessarily followed in other languages, e.g., in Latin and Greek. At any rate, since in our dependency grammar we have agreed to consider prepositions

as elements in three-term structures, we may pose the question of the relative position of the other two elements. Within the prepositional phrase, the ordering of the elements is fixed in Russian and in English: preposition/object. For the sake of convenience, we may represent this combination simply as P. Our inquiry, then, is concerned with the relative ordering of the syntactic governor of P.

At the outset, it should be stressed that in our text, the governors of Ps are severely limited as to part-of-speech. With trivial exception, the governors are verbals (predicates and participles), nouns, and adjectives. The search for the governor can therefore be confined to representatives of these word classes in the clause. In our text, adjectives serve as governor in less than 1% of all occurrences of Ps; our limited experience suggests that, with the exception of strong government, a necessary and sufficient condition for establishing the adjective as governor is its position immediately preceding the P. Accordingly, it is appropriate to deal with the G/P problem in terms of verbals (V) and nouns (N) in the clausal environment of the P.

Six possible orderings of N, V, and P exist:

(1) N/P/V, (2) N/V/P, (3) P/N/V, (4) P/V/N, (5) V/P/N, (6) V/N/P. Of these orderings, only Nos. 1 and 6 present a problem. If the ordering described in Nos. 2 and 4 occurs, our data and the syntactic characteristic known as projectivity tells us that the V is G [7]. In sentences with the sequence N/V/P ("The words/were written/ on the blackboard") or P/V/N ("On the blackboard/were written/ the words"), the P cannot depend on the N. If the ordering described in Nos. 3 and 5 occurs, our experience is that the V is G; to put it differently, in our text, a P does not depend on a noun to its right. This conclusion is based on a sample of prepositional occurrences in the physics text:



Right-Hand Governors of Prepositions

P	Frequency of P	Occurrences of following G	% of Occurrences with following G	Non-Predicate following G	
				Adj.	N
DLYA	1395	486	35%	4	0
DO	104	3	3%	1	0
IZ	280	58	20%	0	0
K	700	34	5%	3	0
NA	1315	212	16%	2	2
O	259	3	1%	1	0
OT	769	4	.5%	0	0
PO	749	114	15%	0	0
S	505	38	7%	0	0
V	<u>1629</u>	<u>288</u>	17%	<u>2</u>	<u>2</u>
	7705	1240		13	4

The two occurrences of a following noun governor for the preposition "V" were special cases, involving larger coordinate structures identifiable by punctuation. The two occurrences with the preposition "NA" were contained in the clauses: NA X OKAZYVAYUT VLIYANIE KAKIE-TO KOMPONENTY = "on x exert an influence certain components" and DAT' NA VOPROS OTVET: = "give to the question the answer:"; in both of these clauses we have instances of strong government, and, in addition, verb/noun phrases that are equivalent to, or transformations of, verbs (exert an influence = to influence, give an answer = to answer). Assuming we have the capability to recognize either type of structure, the occurrences of a following noun governor for a preposition are reduced to zero in our sample.

The extremely small incidence of a right-hand noun governor for Ps is probably characteristic of the

"scientific" prose from which our sample is taken. The incidence may be somewhat higher in other kinds of written discourse, both in Russian and English, and is certainly higher in the spoken language. For example, with the (English) expletive, "there is," the inverted order PVN is possible: "To this room there are three entrances," or "To this question there is no good answer." It seems likely that even in Russian such inversions are extremely rare except in cases of strong government. (The use of such frames to test "strong" versus "weak" government is indicated.) One may apparently discount in written Russian scientific text the possibility that a "weakly governed" or "adjoined" P will precede its noun governor (cf. S RUZH'EM CHELOVEKA UVIDEL = "with the gun the man I saw," or S RUZH'EM UVIDEL CHELOVEKA = "with the gun (I) saw the man.") In any event, we presume in our parsing program the ability to recognize cases of strong government; we merely note here the absence of right-hand noun governors of Ps in the case of weak government or adjoinment.

To summarize: when P, N, and V occur in conjunction, as they often do, there are grounds for utilizing the relative position of these elements as a means of determining the syntactic governor of the preposition: (1) If the P is nearer the V than it is to the N, the V is the governor (PVN and NVP); (2) if the N follows the P, the V is the governor (PNV and VPN). We can then turn to the two remaining sequences of these elements (NPV and VNP), again confining our study to instances of other than strong government.

Concordances for prepositions in the NPV and VNP sequences have not yet been prepared for our complete text. Fairly extensive sampling indicates that the sequence NP occurs in about one-third of all occurrences of Ps, and that in such environments there is a great

tendency for the noun to be the governor. Thus, in one sample of 39 pages of text written by 13 different authors, 1046 Ps were encountered; of these, 342 (33%) occurred in the sequence NPV or VNP. Excluding the few instances of strong government by the predicate, the V was the governor in only 33 instances (9%). The criterion of position could be used to assign the N as correct governor in nine out of ten instances, given the absence of other criteria. (We stress the fact that if a series of Ns precedes the P, the question still remains: which N is governor? See the discussion below.)

An examination of the 33 instances in which the V governed the P in the given structures suggests that morphological criteria are inadequate, i.e., it is not sufficient to determine the part-of-speech of words in context. Leaving aside the question of meaning, it is important to take into account both the function of the prepositional phrase itself and the function of the preceding noun. Here, two facts may be noted: (1) the prepositional phrase itself sometimes served an adverbial function, so that its dependency on the V (rather than on the N) is strongly suggested or required. For example, in the fragment, ETA LINIYA V DEJSTVITEL'NOSTI NE YAVLYAETSYA... = "this line in reality is not," the phrase V DEJSTVITEL'NOSTI = "in reality" serves an adverbial function similar to the function of the adverb DEJSTVITEL'NO = "really." This fact should prohibit its being connected with the preceding noun, LINIYA = "line." Other prepositional phrases serving an adverbial function included V PRINTSIPE = "in principle," V SREDNEM = "on the average," V MINIMUME = "at the minimum," S TOCHNOST'YU = "with an accuracy," and V DVA RAZA = "twice." (2) The preceding N sometimes serves an adverbial function, either as the final element of a prepositional phrase or as an instrumental noun dependent of the verb: USPEKHI BYLI

DOSTIGNUTY V POSLEDNEE VREMYA V IZUCHENII = "successes were achieved recently in the study." Here too, VREMYA in V POSLEDNEE VREMYA = "recently" could be excluded as a potential governor of the P, clearing the way for consideration of the verb as the governor.

It is not clear that an adverbial function can be assigned a priori to certain prepositional phrases, regardless of context. The question requires further study. If these phrases can be treated as fixed combinations with a fixed syntactic function, the 33 exceptions referred to above would be reduced to 15 in our sample, i.e., to 1.4% of the total occurrences of Ps or 4% of the cases of N/P. The following are typical constructions for which no parsing solution is offered: POLE FORMIRUET KARTINU NA EKRANE = "the field forms a picture on the screen," SVET PROPUSKAETSYA CHEREZ TRUBKU S PARAMI NATRIYA = "light is passed through the tube with sodium vapors," EKSPERIMENTOV PROVEDENNYKH V LABORATORII PO VISUALIZATSII = "experiments conducted in the laboratory on the visualization (of)," RASSEYANIE IZUCHALOS' V RABOTE [1] DLYA MISHENEJ = "scattering was studied in paper [1] for targets." Some of these structures are inherently ambiguous; the majority require the application of semantic criteria, including an understanding of the subject matter (i.e., there are constructions that only a physicist can parse confidently and correctly).

The point of the preceding discussion is that constructions difficult or impossible to parse automatically are encountered infrequently in running text. To the writer, this conclusion was unexpected.

The major problem remaining is the selection of the correct noun governor from a series of nouns preceding the P. Constructions of this type (N.../N/P) constitute something less than half of the occurrences of N/P; the first noun in the series is of course modified by a string

of nouns, normally in the genitive case, occasionally in the instrumental or dative case. Assuming the absence of strong government, which noun shall be chosen as governor of the following P?

To this author's knowledge, the only extensive, programmatic solution of this problem is that advanced by Shelimova [8]. In this system, Russian Ps are divided into three main types: those that (in our terminology) cannot depend on a noun, those that can depend only on a noun that is a deverbative, and those that can depend on any noun. The latter group, which includes all the frequent Ps, is divided into seven sub-groups; for each sub-group various criteria are established for specification of the G. These criteria include the deverbative character of members of the preceding noun series, the deverbative character of the dependent of the P, the nominative case of a preceding noun, and the "spatial" significance of preceding Ns or of a V in the sentence. For example, with the preposition "S", the deverbative N in the preceding N series is chosen as the G: PRIMENENIE ETOGO KRITERIYA S OSTOROZHNOST'YU VPOI NE DOPUSTIMO = "the application of this criterion with care is completely admissible." For the preposition "PRI," in the absence of a preceding deverbative N and given a deverbative N object of the P, the predicate is chosen as G: CHITATEL' VSTRETIT MNOGO DRUGIKH PRIMEROV TAKOGO RODA PRI IZUCHENII RAZLICHNYKH OTDELOV MATEMATIKI = "the reader will meet many other examples of this kind in the study of different branches of mathematics."

Shelimova properly disclaims infallibility for her program, remarking occasionally that a given routine will be correct more often than not. In some instances, no solution is obtained. Although the present author is not in agreement with some of the classifications given by Shelimova, and has noted a number of errors that would

result from the application of her program, he agrees that the principles advanced are valid and generally useful. The whole problem, one of the most complicated in automatic parsing, deserves special study. For the present, we can only suggest the following: (i) the volume and complexity of the task requires that automatic methods be used in the compilation of these constructions from written text and in the building of variously sorted concordances; (ii) further knowledge of the syntactic function of word combinations is required, including a better understanding of noun/genitive noun combinations and of the differing functions of prepositional phrases. In other words, this is not a special, isolated problem, but one that waits upon the further accumulation of information about syntax and its relation to meaning. (What, for instance, is the meaning of the fact that certain nouns cannot govern certain prepositions?) It is most doubtful that a satisfactory solution to our problem can be obtained without this further knowledge.

#### 4. CONCLUSIONS

Samplings of parsed Russian text indicate that the relative position of prepositions and their potential syntactic governors is a useful criterion in automatic parsing programs. Assuming that routines exist to account correctly for strong government of prepositions and for the coordination of sentence elements, the following principles can be incorporated in an automatic parsing program:

(1) The search for governors of Ps will be limited to nouns, adjectives, and predicates (including participles) in the clause.

(2) If the P is clause-initial, the governor is the predicate. (No exceptions to this rule were observed.) "Non-initial" elements, for this purpose, include

introductory words and phrases used as sentence modifiers, and clause markers (coordinate and subordinate conjunctions, relative adverbs, etc.).

(3) If the P immediately follows a predicate, a participle, or an adjective, the preceding occurrence is the governor. (No exceptions to this rule were observed.)

(4) If the P immediately follows a noun, this noun (or one of the nouns in a preceding sequence of nouns) is the governor in approximately 90% of such occurrences. The number of exceptions to this rule can be halved if it can be established by other means that an adverbial function is being served by the prepositional phrase itself, or by an immediately preceding prepositional phrase (which terminates with a noun).

The adequacy of these rules will be tested for a larger text sample. No solution is here offered to the problem of choosing the correct noun governor from the potential governors in a preceding sequence of nouns. In general, our experience with running text suggests that the relative position of sentence elements is a much more significant factor in structuring prepositional phrases than is strong government. A strong governor is almost always discoverable by one of the four rules given above; information about the strength or quality of the connection is in some sense redundant. Word order is also a powerful factor in establishing weaker "semantic" connections between the unit of the prepositional phrase and its syntactic governor. The prospects for successful automatic parsing are greatly increased by the strong tendency of writers of "informational" texts to adhere to word order norms.

REFERENCES

1. Hays, D. G., Basic Principles and Technical Variations in Sentence Structure Determination, The RAND Corporation, P-1984, May 1960.
2. Hays, D. G., and T. W. Ziehe, Studies in Machine Translation--10: Russian Sentence-Structure Determination, The RAND Corporation, RM-2538, April 1960.
3. Peshkovskij, A. M., Russkij Sintaksis v Nauchnom Osveshchenii, Izdanie sed'moe, Moscow, 1956, pp. 286-288.
4. Akademiya Nauk SSSR, Institut Russkogo Yazyka, Grammatika Russkogo Yazyka, Izdatel'stvo Akademii Nauk SSSR, Moscow, 1960. T. II, Sintaksis, Chast' Pervaya, p. 27.
5. Akademiya Nauk SSSR, Institut Russkogo Yazyka, Grammatika Russkogo Yazyka, Izdatel'stvo Akademii Nauk SSSR, Moscow, 1960, p. 22.
6. Harper, K. E., Machine Translation of Russian Prepositions, The RAND Corporation, P-1941, May 1960.
7. Hays, D. G., Research Procedures in Machine Translation, The RAND Corporation, RM-2916-PR, December 1961.
8. Shelimova, I. N., "Ustanovlenie sintaksicheskikh svyazej predlozhnykh grupp v russkom yazyke," Lingvisticheskie issledovaniya po mashinnomu perevodu, Izdatel'stvo VINITI, Moscow, 1961. An English translation, "Establishing a Syntactic Relationship of Prepositional Groups in the Russian Language," is available through the U. S. Joint Publications Research Service, Washington, D. C. (JPRS 13173, 27 March 1962, pp. 155-191.